INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

PUBLICATION
INTERNE
N° 1726

A CENTRAL LIMIT THEOREM AND IMPROVED ERROR BOUNDS FOR A HYBRID-MONTE CARLO SEQUENCE WITH APPLICATIONS IN COMPUTATIONAL FINANCE

GIRAY ÖKTEN, BRUNO TUFFIN, VADIM BURAGO

ISSN 1166-8687

IRISA
CAMPUS UNIVERSITAIRE DE BEAULIEU - 35042 RENNES CEDEX - FRANCE

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES
Campus de Beaulieu – 35042 Rennes Cedex – France
Tél. : (33) 02 99 84 71 00 – Fax : (33) 02 99 84 71 71
http://www.irisa.fr

A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance

Giray Ökten*, Bruno Tuffin**, Vadim Burago***

Systèmes communicants
Projet Armor

Publication interne n° 1726 — Juin 2005 — 28 pages

Abstract: In problems of moderate dimensions, the quasi-Monte Carlo method usually provides better estimates than the Monte Carlo method. However, as the dimension of the problem increases, the advantages of the quasi-Monte Carlo method diminish quickly. A remedy for this problem is to use hybrid sequences; sequences that combine pseudorandom and low-discrepancy vectors.

In this paper we discuss a particular hybrid sequence called the mixed sequence. We will provide improved discrepancy bounds for this sequence and prove a central limit theorem for the corresponding estimator. We will also provide numerical results that compare the mixed sequence with the Monte Carlo and randomized quasi-Monte Carlo methods.

Key-words: Simulation, randomized quasi-Monte Carlo, financial options

(Résumé : tsvp)

* okten@math.fsu.edu
** btuffin@irisa.fr
*** burago@tinro.ru

Centre National de la Recherche Scientifique (UMR 6074) Université de Rennes 1 – Insa de Rennes
Institut National de Recherche en Informatique et en Automatique – unité de recherche de Rennes

Un théorème de la limite centrale et des bornes de l'erreur améliorées pour une suite Monte Carlo hybride, avec applications en finance

Résumé : Pour les problèmes de dimension mathématique modérée, la méthode quasi-Monte Carlo fournit habituellement de meilleures estimations que celle de Monte Carlo. Cependant, quand la dimension augmente, les avantages de quasi-Monte Carlo diminuent rapidement. Un remède à ce problème est d'utiliser des suites hybrides, suites qui combinent vecteurs pseudo-aléatoires et vecteurs à discrépance faible.

Dans ce rapport nous discutons d'une suite hybride particulière appelée suite mixte. Nous fournissons des bornes améliorées de la discrépance pour cette suite et donnons un théorème de la limite centrale pour l'estimateur correspondant. Nous fournissons aussi des résultats numériques qui comparent la suite mixte avec Monte Carlo et quasi-Monte Carlo randomisé.

Mots clés : Simulation, quasi-Monte Carlo randomisé, options financières

1 Introduction

In high dimensional problems, quasi-Monte Carlo methods (QMC) start losing their effectiveness over Monte Carlo methods (MC). The dimension above which QMC is no longer competitive depends on the problem in hand. Methods such as Anova decomposition of functions, and concepts such as effective dimension (Caflisch et al [1]) have been used in the past to understand the relationship between the dimension of the function and the accuracy of QMC.

In order to address the potential difficulties of QMC in high dimensions, several authors introduced "hybrid" methods that make use of low-discrepancy sequences in some elaborate way, often combining them with pseudorandom numbers. Examples of such methods are the "mixed" and "scrambled" strategies used by Spanier [1] , the mixed sequence used by Ökten [2, 3], the "renumbering" and "continuation" methods used by Moskowitz [4], and similar numbering techniques used by Coulibaly and Lécot [5], Morokoff and Caflisch [6], and Lécot and Tuffin [7]. The authors of these studies report favorable numerical results when the errors obtained from these hybrid methods are compared with the MC and QMC errors.

In this paper, we will discuss in detail methods that have been named as the mixed method, padding with MC, and padding with randomized QMC (RQMC) [8]. Consider the problem of estimating

$$I = \int_{(0,1)^s} f(x)dx \tag{1}$$

using sums of the form

$$\hat{I} = \frac{1}{N}\sum_{k=1}^{N} f(x^{(k)}) \tag{2}$$

where $x^{(k)}$ are s-dimensional vectors chosen appropriately. If the dimension s is large, and if it is possible to identify a smaller subset of d important variables $\{i_1, ..., i_d\}$, then one has the following options:

1. Sample $\{i_1, ..., i_d\}$ using a d-dimensional QMC sequence, and for the rest of the variables use an $(s-d)$-dimensional MC (pseudorandom) sequence (called the mixed method, or padding QMC by MC);

2. Sample $\{i_1, ..., i_d\}$ using a d-dimensional RQMC sequence, and for the rest of the variables use an $(s-d)$-dimensional MC (pseudorandom) sequence (called the randomized mixed method, or padding RQMC by MC).

Let $x^{(k)} = (q^{(k)}, X^{(k)})$ be an s-dimensional sequence obtained by concatenating the vectors $q^{(k)}$ and $X^{(k)}$. Here $(q^{(k)})_{k\geq 1}$, is a d-dimensional QMC sequence, and $X^{(k)}, k \geq 1$, are independent random variables with the uniform distribution on $(0,1)^{s-d}$. We will call $x^{(k)}$ a mixed sequence. The underlying sequences used in both of the strategies mentioned above are mixed sequences. The first strategy, in computing (2), uses a single mixed sequence to obtain the estimate $\hat{I}$, whereas the second strategy uses independent replications of a mixed sequence, where each replication involves an independent selection of an RQMC sequence, and random vectors $X^{(k)}, k \geq 1$. In our definition of $x^{(k)}$ we took the first d dimensions to be "important" for convenience. The results of the paper are still valid if the important d variables occurred at arbitrary locations. In Section 4, we will discuss these strategies in more detail and present a computational framework that will enable us to compare their effectiveness numerically.

In the next section, we will investigate the discrepancy of the mixed sequence, which is the underlying sequence in the strategies mentioned above. The reason we study the discrepancy is the Koksma-Hlawka inequality, which states that the error, $|I-\hat{I}|$, is bounded by the variation of f (in the sense of Hardy and Krause) multiplied by the discrepancy of the sequence, and thus smaller discrepancy suggests smaller error. The results of this section generalizes the earlier results given in Ökten [2]. In Section 3, we will prove a central limit theorem for the estimator used in strategy 1. And in Section 4 we will present numerical results from computational finance.

2 A Hoeffding-type inequality and discrepancy upper bound

In the following $x^{(k)} = (q^{(k)}, X^{(k)})$ is the kth element of the s-dimensional mixed sequence, where $q^{(k)}$ and $X^{(k)}$ are the deterministic and stochastic components of dimension d and $s-d$. We will write the components of a vector α as $(\alpha_1, ..., \alpha_s)$. Let $\varphi^+(\cdot)$ be a nonnegative nondecreasing function, and $\varphi^-(\cdot)$ be a nonnegative nonincreasing function. We have

$$P\{|Y - E[Y]| \geq \varepsilon\} \leq \frac{E[\varphi^+(Y)]}{\varphi^+(E[Y]+\varepsilon)} + \frac{E[\varphi^-(Y)]}{\varphi^-(E[Y]-\varepsilon)}, \tag{3}$$

for any random variable Y.

Put

$$\varphi^+(x) = e^{\lambda_1 x}, \text{ and } \varphi^-(x) = e^{-\lambda_2 x}, \tag{4}$$

where $\lambda_1, \lambda_2 > 0$, and let

$$Y = \frac{1}{N}\sum_{k=1}^{N} 1_{[0,\alpha)}(x^{(k)}), \text{ where } \alpha = (\alpha', \alpha'').$$

In the above notation, α' is the $d-$dimensional vector that consists of the first d components of the s dimensional vector α. Similarly, we define α'' as the $(s-d)$ dimensional vector that consists of the rest of the components. The interval $[0, \alpha)$ is defined as $\prod_{k=1}^{s}[0, \alpha_k)$.

Observe that $x^{(k)} < \alpha$ iff $q^{(k)} < \alpha'$ and $X^{(k)} < \alpha''$, and hence

$$P\{x^{(k)} < \alpha\} = 1_{[0,\alpha')}(q^{(k)})P\{X^{(k)} < \alpha''\}.$$

Clearly, $P\{X^{(k)} < \alpha''\} = \prod_{k=d+1}^{s} \alpha_k$ which we will simply denote by p. We have

$$E[Y] = \frac{p}{N}\sum_{k=1}^{N} 1_{[0,\alpha')}(q^{(k)}) = \frac{pA}{N}$$

$$Var(Y) = \frac{1}{N^2}\sum_{k=1}^{N} 1_{[0,\alpha')}(q^{(k)})(p-p^2) = \frac{p(1-p)}{N^2}A,$$

where we denote the constant $\sum_{k=1}^{N} 1_{[0,\alpha')}(q^{(k)})$ by A.

From (3) and (4) we get

$$P\{|Y - E[Y]| \geq \varepsilon\} \leq \frac{E[e^{\lambda_1 Y}]}{e^{\lambda_1(E[Y]+\varepsilon)}} + \frac{E[e^{-\lambda_2 Y}]}{e^{-\lambda_2(E[Y]-\varepsilon)}},$$

for any $\lambda_1, \lambda_2 > 0$. Let

$$M_1(\varepsilon) = \inf_{\lambda_1>0} \frac{E[e^{\lambda_1 Y}]}{e^{\lambda_1(E[Y]+\varepsilon)}} \text{ and } M_2(\varepsilon) = \inf_{\lambda_2>0} \frac{E[e^{-\lambda_2 Y}]}{e^{-\lambda_2(E[Y]-\varepsilon)}}.$$

Since λ_1, λ_2 are arbitrary, we have

$$P\{|Y - E[Y]| \geq \varepsilon\} \leq M_1(\varepsilon) + M_2(\varepsilon). \tag{5}$$

Lemma 1 *We have*

$$M_1(\varepsilon) = \exp\left[-AH(\frac{\varepsilon N}{A})\right] \text{ and } M_2(\varepsilon) = \exp\left[-AH(-\frac{\varepsilon N}{A})\right]$$

where

$$H(x) = (1-p)\left(1 - \frac{x}{1-p}\right)\log\left(1 - \frac{x}{1-p}\right) + p\left(1 + \frac{x}{p}\right)\log\left(1 + \frac{x}{p}\right)$$

provided $A > 0$ *and* $0 < \varepsilon < \frac{A}{N}\min\{p, 1-p\}$.

Proof. *We have*

$$\begin{aligned} E[\exp(\lambda Y)] &= \prod_{k=1}^{N} E\left[\exp\left(\frac{\lambda}{N} 1_{[0,\alpha)}(x^{(k)})\right)\right] \\ &= \prod_{k=1}^{N}\left[p 1_{[0,\alpha')}(q^{(k)})(\exp(\lambda/N) - 1) + 1\right], \end{aligned}$$

and since A *is the number of terms where* $1_{[0,\alpha')}(q^{(k)}) = 1$, *the above product simplifies to*

$$E[\exp(\lambda Y)] = (p\exp(\lambda/N) - p + 1)^A.$$

Together with the fact that $E[Y] = pA/N$, *this equation yields*

$$\frac{E[\exp(\lambda Y)]}{\exp(\lambda(E[Y] + \varepsilon)} = (p\exp(\lambda/N) - p + 1)^A \exp(-\lambda(\frac{pA}{N} + \varepsilon)).$$

To find the infimum of this positive quantity, we find the infimum of its logarithm and then take its exponential, i.e.,

$$\begin{aligned} &\inf(p\exp(\lambda/N) - p + 1)^A \exp(-\lambda(\frac{pA}{N} + \varepsilon)) \\ &= \exp\left[\inf\left(A\log(p\exp(\lambda/N) - p + 1) - \lambda(\frac{pA}{N} + \varepsilon)\right)\right] \\ &= \exp\left[N\inf\left(\frac{A}{N}\log(p\exp(\lambda/N) - p + 1) - \frac{\lambda}{N}(\frac{pA}{N} + \varepsilon)\right)\right]. \end{aligned}$$

Consider the function

$$\phi(t) = \frac{A}{N}\log(p\exp(t) - p + 1) - t(\frac{pA}{N} + \varepsilon)$$

where $t = \lambda/N$. *The function* $\phi(t)$ *attains its minimum value at*

$$t_{\min} = \log\left[\frac{p(1-p)A/N - p\varepsilon + \varepsilon}{p(1-p)A/N - p\varepsilon}\right] > 0,$$

assuming that ε is sufficiently small and N is large so that $A > 0$ and $\frac{(1-p)A}{N} > \varepsilon$. Then

$$\phi(t_{\min}) = -\left(\frac{A}{N} - \varepsilon\right)\log\left(\frac{A/N - \varepsilon}{1-p}\right) - \varepsilon\log\left(\frac{\varepsilon}{p}\right) + \frac{A}{n}\log(A/N).$$

Therefore we have

$$M_1(\varepsilon) = \inf_{\lambda>0}\frac{E[\exp(\lambda Y)]}{\exp(\lambda(E[Y]+\varepsilon))} = \exp(N\phi(t_{\min}))$$

and it can be shown that

$$N\phi(t_{\min}) = -AH(\varepsilon N/A)$$

where $H(\cdot)$ is the function given in the statement of Lemma 1. The expression for $M_2(\varepsilon)$ is obtained similarly. ■

From (5) and Lemma 1, we have

$$\begin{aligned} P\{|Y - E[Y]| \geq \varepsilon\} &\leq \exp\left[-AH\left(\frac{\varepsilon N}{A}\right)\right] + \exp\left[-AH\left(-\frac{\varepsilon N}{A}\right)\right] \\ &\leq 2\exp\left[-A\min\{H(\varepsilon N/A), H(-\varepsilon N/A)\}\right]. \end{aligned}$$

Since

$$H''(x) = \frac{1}{(x+p)(1-p-x)} \geq K = \min_{x,p}\frac{1}{(x+p)(1-p-x)} > 0$$

then $H(x)$ is strongly convex and so for any admissible x and x_0

$$H(x) \geq H(x_0) + H'(x_0)(x - x_0) + \frac{K}{2}(x - x_0)^2.$$

Choosing $x_0 = 0$ and checking up the bound $K = 4$ we obtain that $H(x) \geq 2x^2$, hence $\min\{H(\varepsilon N/A), H(-\varepsilon N/A)\} \geq 2\varepsilon^2 N^2/A^2$ and thus

$$P\{|Y - E[Y]| \geq \varepsilon\} \leq 2\exp\left[-2\varepsilon^2 N^2/A\right]. \tag{6}$$

Consider the local discrepancy random variable

$$g(\alpha) = \frac{1}{N}\sum_{k=1}^{N} 1_{[0,\alpha)}(x^{(k)}) - \prod_{k=1}^{s}\alpha_k = Y - \prod_{k=1}^{s}\alpha_k.$$

Taking expectations, we get $E[g(\alpha)] = E[Y] - \prod_{k=1}^{s} \alpha_k$, and subtracting the equations we obtain

$$g(\alpha) - E[g(\alpha)] = Y - E[Y]. \tag{7}$$

We also note the following inequality

$$\begin{aligned} D_N^*(q^{(k)}) &= \sup_{\alpha' \in (0,1)^d} \left| \frac{1}{N} \sum_{k=1}^{N} 1_{[0,\alpha')}(q^{(k)}) - \prod_{k=1}^{d} \alpha_k \right| \\ &\Rightarrow \frac{A}{N} - \prod_{k=1}^{d} \alpha_k \le D_N^*(q^{(k)}) \\ &\Rightarrow \frac{A}{N} \le D_N^*(q^{(k)}) + \prod_{k=1}^{d} \alpha_k \le D_N^*(q^{(k)}) + 1 \end{aligned}$$

for any α'. Then we get

$$\frac{\varepsilon^2 N^2}{A} \ge \frac{\varepsilon^2 N^2}{N\left(D_N^*(q^{(k)}) + 1\right)} = \frac{\varepsilon^2 N}{D_N^*(q^{(k)}) + 1}. \tag{8}$$

From (6), (7), and (8), we obtain

$$P\{|g(\alpha) - E[g(\alpha)]| \ge \varepsilon\} \le 2 \exp\left[\frac{-2\varepsilon^2 N}{D_N^*(q^{(k)}) + 1}\right].$$

It can be shown that

$$|g(\alpha) - E[g(\alpha)]| < \varepsilon \Rightarrow |g(\alpha)| < \varepsilon + D_N^*(q^{(k)})$$

and thus

$$P\{|g(\alpha)| < \varepsilon + D_N^*(q^{(k)})\} \ge P\{|g(\alpha) - E[g(\alpha)]| < \varepsilon\} \ge 1 - 2\exp\left[\frac{-2\varepsilon^2 N}{D_N^*(q^{(k)}) + 1}\right].$$

In other words, $|g(\alpha)| < \varepsilon + D_N^*(q^{(k)})$ with probability greater than or equal to $1 - 2\exp\left[\frac{-2\varepsilon^2 N}{D_N^*(q^{(k)})+1}\right]$, for any α. Since the upper bound for $|g(\alpha)|$, and the probability, do not depend on α, and since $\sup_\alpha |g(\alpha)| = D_N^*(x^{(k)})$, we have proved,

Theorem 2 *Let $x^{(k)} = (q^{(k)}, X^{(k)})$ be an s-dimensional mixed sequence, where $q^{(k)}$ is a d-dimensional deterministic sequence, and $X^{(k)}$ is a random variable with the uniform distribution on $(0,1)^{s-d}$. Then for any $\varepsilon > 0$*

$$D_N^*(x^{(k)}) < \varepsilon + D_N^*(q^{(k)}),$$

with probability greater than or equal to

$$1 - 2\exp\left[\frac{-2\varepsilon^2 N}{D_N^*(q^{(k)}) + 1}\right].$$

Corollary 3 *Put $\varepsilon := (\varepsilon_N) = (N^{-a/2}), 0 < a < 1$, in the above theorem, and let $\{q^{(k)}\}_{k=1}^{\infty}$ be a low-discrepancy sequence with $D_N^*(q^{(k)}) \leq c_d \frac{(\log N)^d}{N} + O\left(\frac{(\log N)^{d-1}}{N}\right)$. Then the discrepancy of the mixed sequence satisfies*

$$D_N^*(x^{(k)}) < \frac{1}{N^{a/2}} + c_d \frac{(\log N)^d}{N} + O\left(\frac{(\log N)^{d-1}}{N}\right), \tag{9}$$

with probability greater than or equal to

$$1 - 2\exp\left[-2\left(c_d \frac{(\log N)^d}{N^{2-a}} + O\left(\frac{(\log N)^{d-1}}{N^{2-a}}\right) + \frac{1}{N^{1-a}}\right)^{-1}\right]. \tag{10}$$

The best values for $c_d, 2 \leq d \leq 20$, are calculated by Niederreiter for the $(t,s)-$sequences constructed by him in [9]. Omitting the lower order terms, let $A_1 = c_s N^{-1}(\log N)^s$ be the upper bound for the discrepancy of the s-dimensional low-discrepancy sequence, and $A_2 = N^{-a/2} + c_d N^{-1}(\log N)^d$ be the probabilistic upper bound (9) for the mixed(s,d) sequence. Similarly, the lower bound (10) for the probability that the discrepancy bound is satisfied is

$$A_3 = 1 - 2\exp\left[-2\left(c_d \frac{(\log N)^d}{N^{2-a}} + \frac{1}{N^{1-a}}\right)^{-1}\right].$$

In the following table, we compute these bounds using two-digit rounding arithmetic, when $N = 10^7, a = 0.9, d = s/2$, and $s = 4, 6, ..., 20$.

s	A_1	A_2	A_3
4	5.8×10^{-4}	7.1×10^{-4}	1
6	3.3×10^{-2}	7.6×10^{-4}	1
8	1.4	1.3×10^{-3}	1
10	5.1×10	3.4×10^{-3}	1
12	1.7×10^{3}	3.3×10^{-2}	1
14	1.7×10^{5}	1.2×10^{-1}	1
16	1.6×10^{6}	1.4	9.7×10^{-1}
18	4.6×10^{7}	4.4	6.8×10^{-1}
20	4.6×10^{9}	5.1×10	-6.5×10^{-1}

We note that when $s = 20$, the lower bound for the probability becomes negative, and therefore useless. Upto dimension $s = 14$, the discrepancy of mixed sequences satisfy the upper bounds given by A_2 with probability 1, and these upper bounds are smaller than the corresponding upper bounds for the s-dimensional low-discrepancy sequences in all cases except $s = 4$, with factors of improvement as high as 10^6.

3 A central limit theorem for the mixed method

The problem we are interested in is the estimation of the integral of a bounded function over the s-dimensional hypercube

$$I = \int_{[0,1]^s} f(x)dx,$$

using the estimator

$$\theta_m = \frac{1}{N}\sum_{k=1}^{N} f(x^{(k)})$$

where $\{x^{(k)}\}_{k=1}^{\infty}$ is the $s-$dimensional mixed sequence

$$x^{(k)} = (q_1^{(k)}, \cdots, q_d^{(k)}, X_{d+1}^{(k)}, \cdots,, X_s^{(k)}).$$

Define the random variables

$$Y_k = f\left(q_1^{(k)}, \cdots, q_d^{(k)}, X_{d+1}^{(k)}, \cdots,, X_s^{(k)}\right),$$

let $\mu_k = E[Y_k]$ and $\sigma_k^2 = Var(Y_k)$ and

$$s_N^2 = Var(\theta_m)N^2 = \sigma_1^2 + ... + \sigma_N^2.$$

We will next prove a central limit theorem stating that, (1) The estimator θ_m is asymptotically normally distributed; (2) Its asymptotic variance is theoretically known; (3) The estimator has a smaller variance than the MC method asymptotically.

Theorem 4 *Assume that f is bounded over $[0,1]^s$ and the functions*

$$g(x_1,...,x_d) = \int_{[0,1]^{s-d}} f(x_1,...,x_d,X_{d+1},...,X_s)^2 dX_{d+1}...dX_s$$
$$h(x_1,...,x_d) = \left(\int_{[0,1]^{s-d}} f(x_1,...,x_d,X_{d+1},...,X_s)^2 dX_{d+1}...dX_s\right)^2$$

are of bounded variation in the sense of Hardy and Krause (sufficient condition for convergences $\frac{1}{N}\sum_{k=1}^N g(q_1^{(k)},\cdots,q_d^{(k)}) \to \int_{[0,1]^d} f(x)^2dx$ and $\frac{1}{N}\sum_{k=1}^N h(q_1^{(k)},\cdots,q_d^{(k)}) \to \int_{[0,1]^d} h(y)dy = \int_{[0,1]^d}\left(\int_{[0,1]^{s-d}} f(y,x)dx\right)^2 dy$). Then

1. *The distribution of the normalized sum*

$$\frac{\sum_{k=1}^N Y_k - \sum_{k=1}^N \mu_k}{s_N}$$

   tends to the standard normal distribution.

2. *We have*

$$s_N^2/N \to L = \int_{[0,1]^s} f(x)^2dx - \int_{[0,1]^d}\left(\int_{[0,1]^{s-d}} f(y,x)dx\right)^2 dy;$$

3. *The mixed strategy always yields a reduction in the standard MC variance, with the reduction given by*

$$\frac{\int_{[0,1]^s} f(x)^2dx - \int_{[0,1]^d}\left(\int_{[0,1]^{s-d}} f(y,x)dx\right)^2 dy}{\int_{[0,1]^s} f(x)^2dx - \left(\int_{[0,1]^s} f(x)dx\right)^2} \leq 1.$$

Proof. *The variance of Y_k is*

$$\sigma_k^2 = \int_{[0,1]^{s-d}} (f(q_1^{(k)}, \cdots, q_d^{(k)}, X_{d+1}, \cdots, X_s))^2 dX_{d+1} \cdots dX_s - \left(\int_{[0,1]^{s-d}} f(q_1^{(k)}, \cdots, q_d^{(k)}, X_{d+1}, \cdots, X_s) dX_{d+1} \cdots dX_s\right)^2$$

Since f is bounded, Y_n are also bounded and, from a standard result (see Feller [10]), it suffices to show that $s_N \to \infty$ when $N \to \infty$ to verify the Lindeberg condition that ensures a central limit theorem for independent but non-identical random variables. But, from the Koksma-Hlawka theorem (see for instance [9]), we have

$$\frac{1}{N} \sum_{k=1}^{N} g(q_1^{(k)}, \cdots, q_d^{(k)}) \to \int_{[0,1]^d} f(x)^2 dx$$

and

$$\frac{1}{N} \sum_{k=1}^{N} h(q_1^{(k)}, \cdots, q_d^{(k)}) \to \int_{[0,1]^d} h(y) dy = \int_{[0,1]^d} \left(\int_{[0,1]^{s-d}} f(y,x) dx\right)^2 dy,$$

proving Claim 2. The Lindeberg condition is satisfied and we get the central limit theorem of Claim 1. For the last claim, we note that $s_N^2/N \to \int_{[0,1]^s} f(x)^2 dx - \int_{[0,1]^d} \left(\int_{[0,1]^{s-d}} f(y,x)dx\right)^2 dy$ as $N \to \infty$ whereas $\sigma^2 = \int_{[0,1]^s} f(x)^2 dx - \left(\int_{[0,1]^s} f(x)dx\right)^2$ is the variance of $f(X)$ for X uniformly distributed over $(0,1)^s$. The fact that we always get a variance reduction comes from

$$\int_{[0,1]^d} \left(\int_{[0,1]^{s-d}} f(y,x) dx\right)^2 dy > \left(\int_{[0,1]^d} \int_{[0,1]^{s-d}} f(y,x) dx dy\right)^2$$

(special case of the Cauchy-Schwarz inequality). ■

Remark 5 *It is important to note that the theorem is valid as long as the deterministic sequence used in the definition of the estimator θ_m is uniformly distributed modulo one. In particular, if we choose the sequence to be a low-discrepancy sequence, its faster convergence rate will help reduce the bias of the estimator, and increase the convergence rate of the variance to its asymptotic value. Both of these observations follow from the Koksma-Hlawka inequality.*

Currently we do not know a practical and efficient way of estimating s_N. An upper bound for s_N, however, can be found using the variance of the MC estimator. Indeed, let us assume that the d-dimensional functions f, f^2 are of bounded variation. Using this fact, and the fact that the discrepancy of the first N points of the sequence $(q_1^{(k)}, ..., q_d^{(k)}, X_{d+1}^{(k)}, ..., X_s^{(k)})_k$ tends almost surely to zero when $N \to \infty$ (since it is uniformly distributed over $[0,1)^s$), we obtain

$$\frac{1}{N}\sum_{k=1}^{N} f^2(q_1^{(k)}, \cdots, q_d^{(k)}, X_{d+1}^{(k)}, \cdots,, X_s^{(k)}) \to \int_{[0,1]^s} f^2(x)dx$$

$$\frac{1}{N}\sum_{k=1}^{N} f(q_1^{(k)}, \cdots, q_d^{(k)}, X_{d+1}^{(k)}, \cdots,, X_s^{(k)}) \to \int_{[0,1]^s} f(x)dx$$

and thus

$$\frac{1}{N}\sum_{k=1}^{N} f^2(q_1^{(k)}, ..., q_d^{(k)}, X_{d+1}^{(k)}, ..., X_s^{(k)}) - \left(\frac{1}{N}\sum_{k=1}^{N} f(q_1^{(k)}, ..., q_d^{(k)}, X_{d+1}^{(k)}, ..., X_s^{(k)})\right)^2 \to \sigma^2$$

almost surely as $N \to \infty$.

4 Randomization and numerical results

4.1 Randomization, estimators and efficiency

In this section we will compare the mixed method with MC and randomized mixed (Rmixed) methods numerically, when they are applied to problems from security pricing. For simplicity, we define our estimators in the context of numerical quadrature; they are extended easily to the more complicated problem from finance. To this end, consider the problem of computing

$$I = \int_{[0,1]^s} f(x)dx.$$

Let $X^{(k)}, k = 1, ...$ be a sequence of i.i.d random variables with distribution $U(0,1)^s$, $X_i^{(k)}, i = 1, ..., ; k = 1, ...,$ be a sequence of i.i.d random variables with distribution $U(0,1)$, $x^{(k)} = (q_1^{(k)}, \cdots, q_d^{(k)}, X_{d+1}^{(k)}, \cdots, X_s^{(k)})$ be the kth element of an s-dimensional mixed sequence with a d-dimensional deterministic component, and let

$u^{(k,i)}$ be the kth element of the ith realization of a mixed sequence whose deterministic component is the ith realization of a d-dimensional RQMC sequence, and the remaining $(s-d)$ components are sampled from $U(0,1)^{s-d}$. We then define estimators (earlier discussed in Introduction):

$$\theta = \frac{1}{NM}\sum_{k=1}^{NM} f(X^{(k)}) \text{ - MC}$$

$$\theta_{mixed} = \frac{1}{NM}\sum_{k=1}^{NM} f(x^{(k)}) \text{ - Mixed (padding QMC by MC)}$$

$$\theta_{Rmixed} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{N}\sum_{k=1}^{N} f(u^{(k,i)})\right) \text{ - Randomized mixed (padding RQMC by MC)}$$

Note that θ_{mixed} is a biased estimator. We want to know how the bias and standard deviation of θ_{mixed} compare with the standard deviations of the unbiased estimators θ and θ_{Rmixed}. Here is one interpretation of the estimators θ_{mixed} and θ_{Rmixed}: θ_{mixed} goes NM "deep" in one realization of the underlying sequence, whereas θ_{Rmixed} goes N "deep" and averages over M realizations of the sequence. Also note that if we take $d = s$ in θ_{Rmixed} (no padding) we obtain the RQMC estimator. In our numerical results we will also compare the methods based on padding with the RQMC estimator.

In the numerical examples, we will consider two implementations of θ_{Rmixed}. One will use the scrambled (t,d) sequences of Owen [11], and the other will use the linear scrambling approach of Matoušek [12, 13]. Both scrambling methods are applied to a $(0,d)$-sequence in base p with p smallest prime number larger than or equal to d. Our main concern is the behavior of the error for moderate sample sizes and how expensive it is to generate the estimates, and thus the existing asymptotical results on the variance of RQMC methods (see [14] and the references mentioned) are not useful to us. Instead we will compare the efficiency of these methods numerically. We define the efficiency $\varepsilon(\theta)$ of an estimator θ as

$$\varepsilon(\theta) = \left(\left(Var(\theta) + (E[(\theta - I)])^2\right) t\right)^{-1}$$

where t is the complexity of the computation. We will estimate $\varepsilon(\theta)$ as follows: t will be taken as the computation time, $E[(\theta - I)]$ will be taken as the computed bias for the θ_{mixed} estimator (in our examples we will know the true answer so that bias can be computed), and $Var(\theta)$ will be the sample variance. For the

MC and Rmixed methods, the variance is estimated like in usual MC methods from the respectively NM and M independent random variables. The variance of the mixed sequence cannot be computed directly (we can only find an upper bound as discussed in the previous section). Instead, we estimate the variance by computing the sample variance of 100 independent replications (i.e., independent uniform random coordinates between the $(d+1)$st and the sth coordinates, the first d determined by the low-discrepancy sequence).

4.2 Pricing of financial securities

Here we consider a problem from computational finance: pricing of geometric Asian options. The price of these options can be computed exactly, however, a close relative, arithmetic Asian options, do not have exact pricing formulas. In simulation, we generate a sequence of asset prices $S_0, S_1, ..., S_K$ that are subject to an Ito process $dS = \mu S dt + \sigma S dX$, where t is time, μ and σ are the drift and volatility of the underlying respectively, and $X = (X(t))_t$ is a standard Brownian motion. The payoff function is defined as $h(S_0, S_1, ..., S_K) = \max(G(S_0, S_1, ..., S_K) - E, 0)$, where $G(S_0, S_1, ..., S_K) = \left(\prod_{i=0}^{K} S_i\right)^{1/(K+1)}$ is the geometric average of the asset prices, and E is the strike price. The price of the option is the expected value $\mathbf{E}\left[e^{-rT} h(S_0, S_1, ..., S_K)\right]$, which is estimated by simulation. In this expression T refers to the expiration time: this is the time when we observe the final price S_K. Details on geometric options, including the exact pricing formula can be found in [15].

We estimated the option price using MC, mixed, and Rmixed methods. In this problem K corresponds to the dimension of the problem (which was denoted by s in the previous sections), and in the first numerical examples K is taken to be 256. The dimension of the deterministic part of the mixed sequence is taken to be $d = 32$. The other constants are: $\mu = 0.1$, $\sigma = 0.1$, $T = 128$, $E = 5$ and $S_0 = 500$, leading to an exact price of 0.76561. The Brownian bridge construction [16] is first used to solve the model, so that most of the variance is concentrated in the first coordinates (even if it is not always the case, see [17]). Recall that the Brownian bridge formula assumes in its simplest implementation that K is a power of 2. From S_0, S_K is first computed, then $S_{K/2}$, $S_{K/4}$, $S_{3K/4}$, $S_{K/8}$, $S_{3K/8}$, $S_{5K/8}$, $S_{7K/8}$ and so on (see [16] for details). Figure 1 displays the results when the number of points NM increases (M is fixed at 100, we only increase N).

We plot confidence interval width (CI width), computation time, bias for the mixed method, and the efficiency in Figure 1. The first three plots give us specific

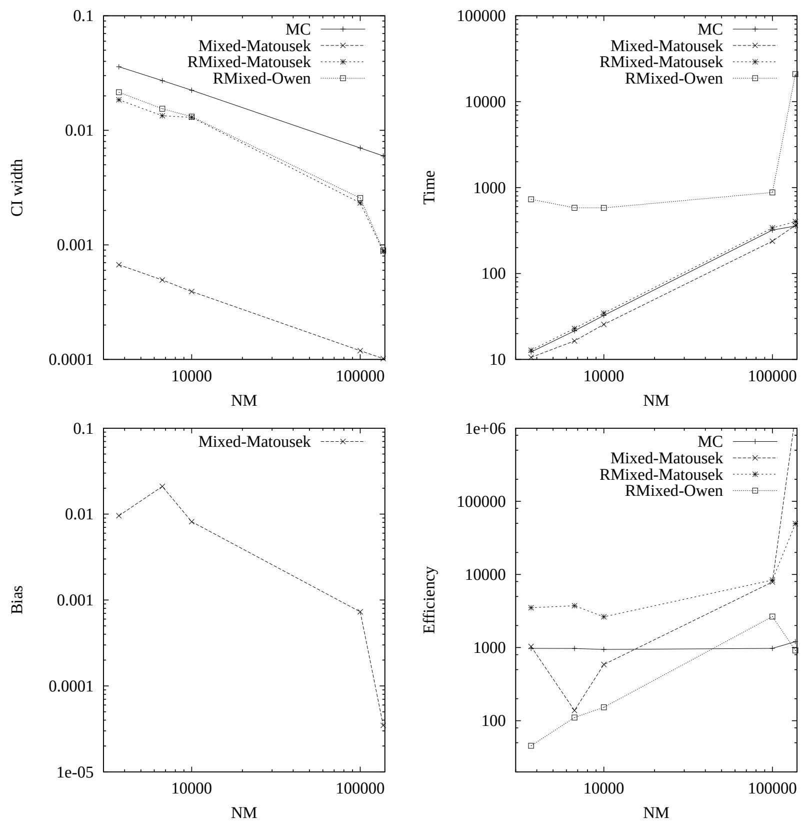

Figure 1: Pricing an asian option in dimension 256 using a 32-dimensional low discrepancy sequence and the Brownian bridge implementation

information about each method, and the last plot for efficiency shows the overall effectiveness of the methods. Among other things we notice the high execution time for the Rmixed-Owen, which is expected, and the way the error for the mixed method is broken into two components as bias and CI width. Overall, Rmixed-Matoušek has the best efficiency ($d = 32$) with an average improvement factor of 4.5 in efficiency over MC. The efficiency of the mixed method is between MC and Rmixed-Owen for the first three samples, and then it gets better, giving the best efficiency for the last sample size.

We next try different values for d, using the Matoušek implementation. Figure 2 compares the results for the case of Rmixed-Matoušek with the above inputs but for $d = 32$, $d = 64$ and, $d = 256$ (which corresponds to the traditional RQMC method - no padding).Note that $d = 32$ gives better efficiency than $d = 256$ (RQMC) for all except one sample size. When $N = 100,000$, the improvement is about a factor of 8.5.

How do these results change if Brownian bridge is not used? Figure 3 solves the same problem and uses the same methods as Figure 1 (except that we ignore the mixed method) without the Brownian bridge implementation. As before, Rmixed-Matoušek has the best efficiency ($d = 32$), but the improvement over MC is approximately a factor of 1.3, which is a smaller improvement than the case when Brownian bridge was employed.

Figure 4 compares different values for d like Figure 2, but without the Brownian bridge implementation. Comparing these two figures we make an interesting observation: When there is no Brownian bridge, the efficiency of RQMC-Matoušek is pretty bad compared to Rmixed methods for smaller sample sizes. However, for larger sample sizes, the efficiencies get closer. If Brownian bridge is used, than exactly the opposite seems to be true; efficiencies are closer for smaller samples, and farther apart for larger samples.

Comparing the plots for CI width in Figure 3 & Figure 1, and Figure 4 & Figure 2 also show that the Brownian bridge implementation lowers the variance for Rmixed and RQMC methods, but not for the MC method.

We now increase the dimension of the problem to $K = 1024$, and compare the efficiency of Rmixed-Matoušek ($d = 32$) with full scrambling, RQMC-Matoušek ($d = 1024$). Figure 5 shows that when Brownian bridge is used the Rmixed-Matoušek ($d = 32$) method has a much better efficiency than the full RQMC-Matoušek, by an average factor of 10, although there is quite a bit of variation. When Brownian bridge is not used, Rmixed-Matoušek has better efficiency for all except one sample size. We also considered large samples and simulated this problem upto $N = 10^7$.

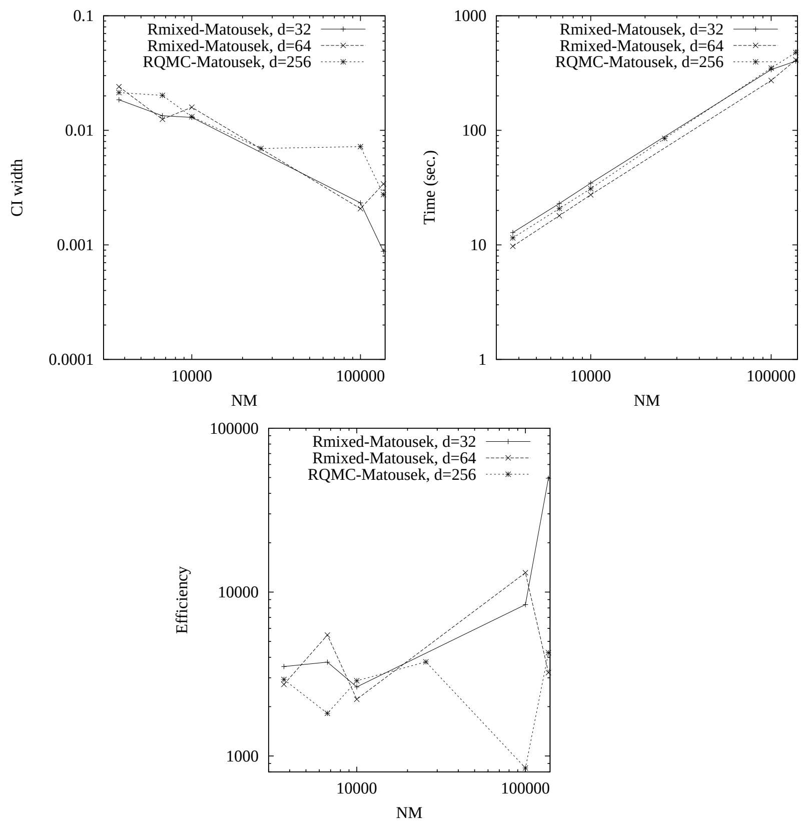

Figure 2: Pricing an asian option in dimension 256 using Rmixed-Matoušek scrambling and different values for d with the Brownian bridge implementation

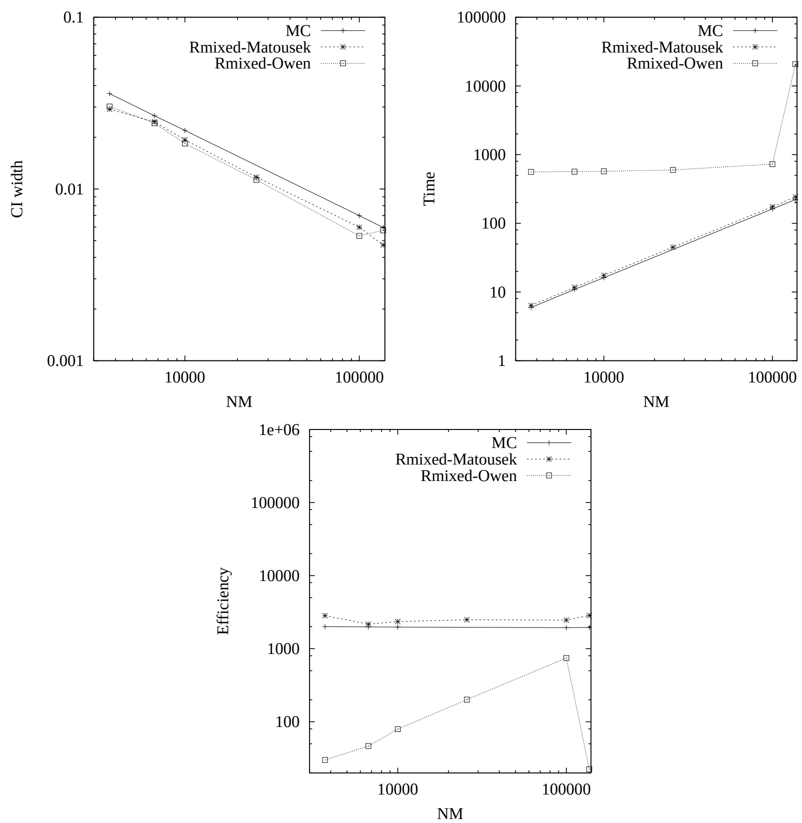

Figure 3: Pricing an asian option in dimension $K = 256$ and $d = 32$, without the Brownian bridge implementation

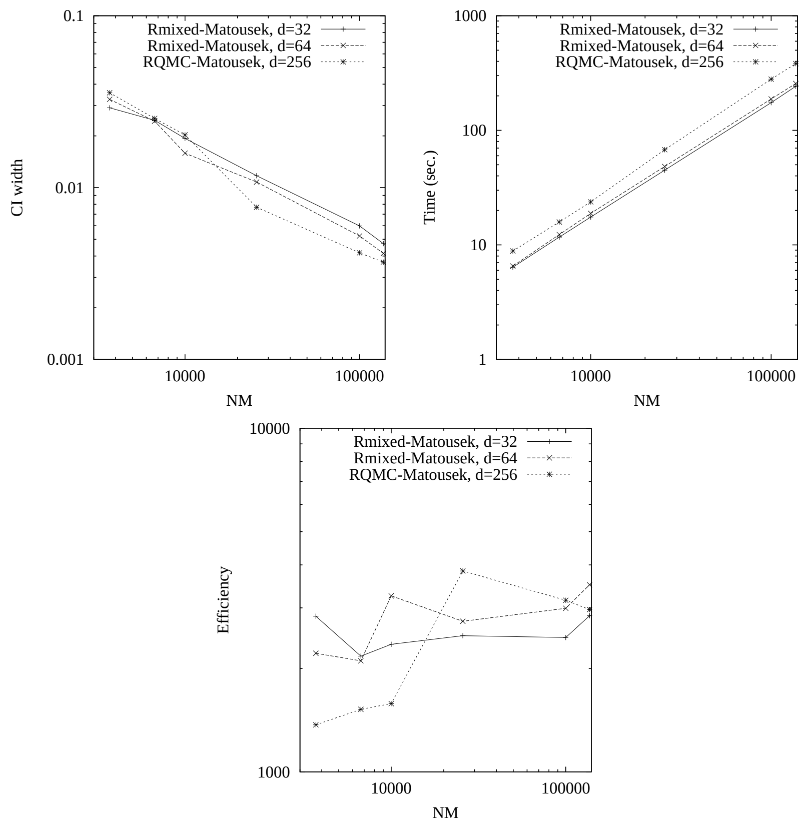

Figure 4: Pricing of an asian option in dimension 256 using Rmixed-Matoušek and different values for d without the Brownian bridge implementation

The efficiency of Rmixed-Matoušek ($d = 32$) gets even better with a wider margin than RQMC-Matoušek as sample size grows, in the case of Brownian bridge implementation. However, if Brownian bridge is not used, RQMC-Matoušek efficiency gets slightly better.

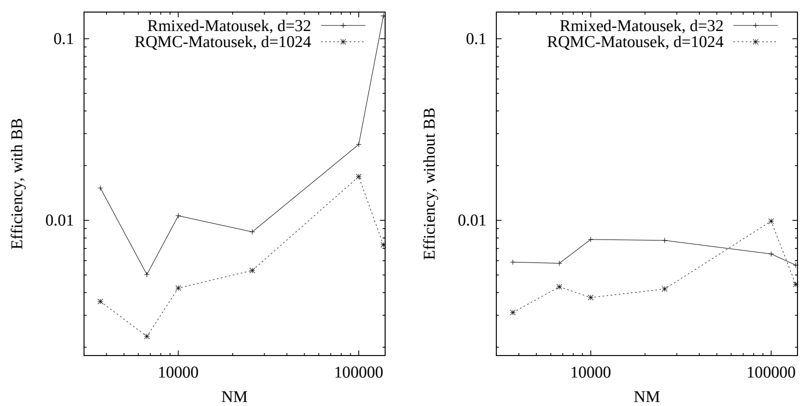

Figure 5: Pricing of an asian option in dimension 1024 using Rmixed-Matoušek with $d = 32$ and RQMC, $d = 1024$. The figure on the left is with the Brownian bridge implementation, and the figure on the right is without the Brownian bridge implementation

Our second example is pricing of digital options. We assume the stock price follows the geometric Brownian motion model as in the Asian option example. The payoff function is

$$h(S_1, ..., S_K) = \frac{1}{K} \sum_{i=1}^{K} (S_i - S_{i-1})^0_+ \; S_j,$$

where $(x)^0_+$ is equal to 1 if $x > 0$; otherwise it is 0. These options were considered by Papageorgiou [17] who showed that the Brownian bridge implementation consistently performed worse than the standard implementation. We therefore do not consider the Brownian bridge implementation in this example.

We start with a 256 dimensional problem and compare Rmixed-Matoušek methods ($d = 32$ and $d = 64$) with the full RQMC-Matoušek implementation. Examining

Figure 6, we make a similar observation we had earlier: The efficiency of RQMC-Matoušek is worse initially than the Rmixed methods, but as the sample size gets larger the efficiencies get closer.

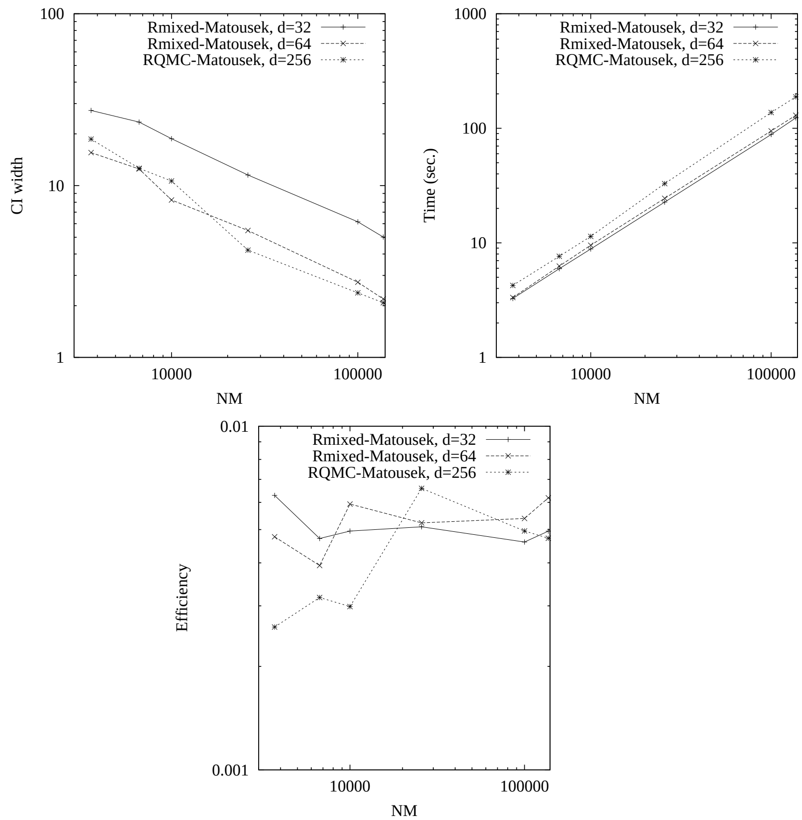

Figure 6: Pricing of a digital option in dimension 256 using Rmixed-Matoušek with $d = 32$, $d = 64$ and RQMC-Matoušek, with $d = 256$.

We now investigate how the biased mixed estimator compares with the others. In Figure 7, we plot the CI width, time, bias, and efficiency when the methods

MC, Mixed-Matoušek ($d = 32$), Rmixed-Matoušek ($d = 32$), and RQMC are used. Perhaps surprisingly, the mixed method gives the best efficiency for all except two sample sizes. Rmixed-Matoušek ($d = 32$) comes second in overall efficiency. Both methods outperform MC consistently, and RQMC efficiency gets close to the mixed and Rmixed methods for large samples.

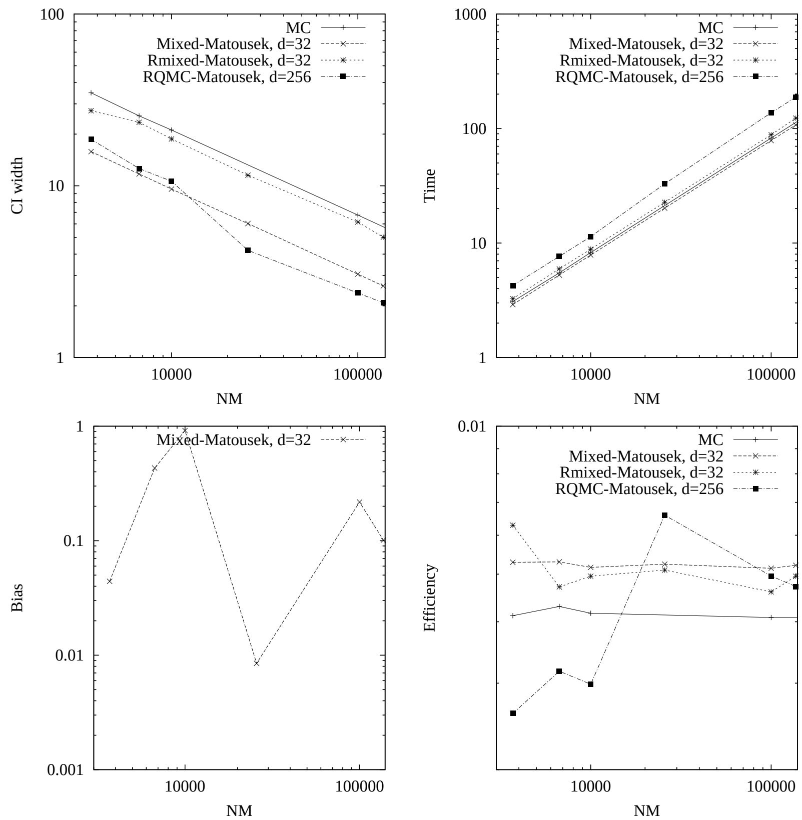

Figure 7: Pricing of a digital option in dimension 256 using MC, Rmixed-Matoušek with $d = 32$, and RQMC-Matoušek, with $d = 256$.

How do these results change if the dimension of the deterministic part of the mixed sequence is increased to 64? In Figure 8, we see that the efficiency of the

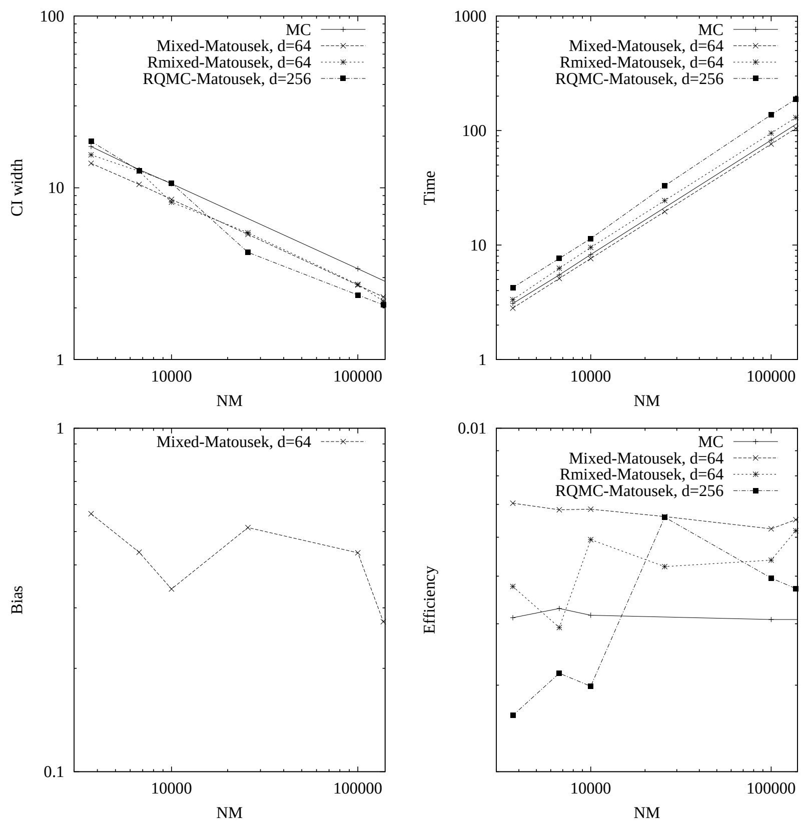

Figure 8: Pricing of a digital option in dimension 256 using MC, Rmixed-Matoušek with $d = 64$, and RQMC-Matoušek, with $d = 256$.

mixed method gets even better: Now the mixed-Matoušek ($d = 64$) efficiency is better than the other methods for all sample sizes but one. The efficiency of mixed-Matoušek ($d = 64$) is about a factor of 1.3 (meaning 30%) better than MC. An approximate figure of merit is harder to come up with due to high oscillations in the efficiency of RQMC-Matoušek and Rmixed-Matoušek ($d = 64$), however, especially for smaller sample sizes, the improvement is pretty noteworthy.

Finally, we look at the efficiency when the dimension is increased to $K = 1024$, and $d = 128$. The mixed-Matoušek has better efficiency than all of the other methods for all except two sample sizes. These results are consistent with the previous ones.

5 Conclusions

In this paper, we studied the mixed method for high-dimensional integration, where the first coordinates are sampled using a QMC sequence and the remaining ones are sampled by MC. The method was known to give good experimental results, but little was known theoretically about the approximation error. We proved an upper bound for the discrepancy of the mixed sequence improving the earlier results of Ökten [2]. Next, we obtained a central limit theorem that enables the use of confidence intervals for the integral. We then discussed numerical results when the mixed method and its randomized versions were applied to problems from option pricing. Our numerical investigations suggest that the mixed method (padding QMC with MC) and its randomized version, the Rmixed method (padding RQMC with MC), can significantly improve efficiency in high dimensional problems for especially moderate sample sizes. Although we see improvements with and without the Brownian bridge implementation, the use of Brownian bridge magnified the factors of improvement in the Asian option example. We also observed that the biased mixed method has the potential of outperforming its mixed version as well as the full RQMC strategy in terms of efficiency. This happens when the bias is small compared to the variance, and there is significant gain in computation time.

References

[1] J. Spanier, Quasi-Monte Carlo Methods for Particle Transport Problems, in: H. Niederreiter, P. J.-S. Shiue (Eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Vol. 106 of Lecture Notes in Statistics, Springer, 1995, pp. 121–146.

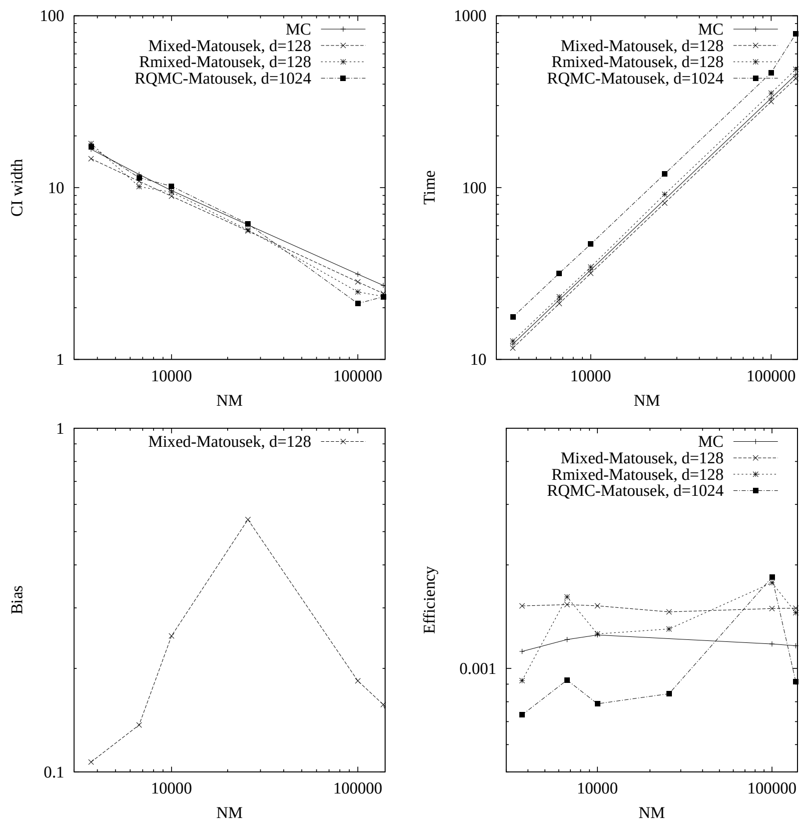

Figure 9: Pricing of a digital option in dimension 1024 using MC, Rmixed-Matoušek with $d = 128$, and RQMC-Matoušek, with $d = 1024$.

[2] G. Ökten, A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications, Monte Carlo Methods and Applications 2 (4) (1996) 255–270.

[3] G. Ökten, High dimensional simulation, Mathematics and Computers in Simulation 55 (2001) 215–222.

[4] B. Moskowitz, Quasirandom diffusion Monte Carlo, in: H. Niederreiter, P. J.-S. Shiue (Eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Vol. 106 of Lecture Notes in Statistics, Springer, 1995, pp. 278–298.

[5] I. Coulibaly, C. Lécot, Monte Carlo and quasi-Monte Carlo algorithms for a linear integro-differential equation, in: G. L. H. Niederreiter, P. Hellekalek, P. Zinterhof (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 1996, Vol. 127 of Lecture Notes in Statistics, Springer Verlag, 1998, pp. 176–188.

[6] W. J. Morokoff, R. E. Caflisch, A Quasi-Monte Carlo Approach to Particle Simulation of the Heat Equation, SIAM Journal of Numerical Analysis 30 (6) (1993) 1558–1573.

[7] C. Lécot, B. Tuffin, Quasi-Monte Carlo Methods for Estimating Transient Measures of Discrete Time Markov Chains, in: Monte Carlo and Quasi-Monte Carlo Methods 2002. Proceedings of a Conference held at the National University of Singapore, Republic of Singapore, November 25-28, 2002, Springer Verlag, 2004, pp. 329–344.

[8] A. Owen, Monte Carlo, quasi-Monte Carlo, and randomized quasi-Monte Carlo, in: H. Niederreiter, J. Spanier (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 1998, Lecture Notes in Statistics, Springer, 2000, pp. 86–97.

[9] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, CBMS-SIAM 63, Philadelphia, 1992.

[10] W. Feller, An Introduction to Probability Theory and its Applications, Vol. II, John Wiley and Sons, 1966, Second edition.

[11] A. B. Owen, Randomly permuted (t, m, s)-nets and (t, s)-sequences, in: H. Niederreiter, P. J.-S. Shiue (Eds.), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Vol. 106 of Lecture Notes in Statistics, Springer, 1995, pp. 299–315.

[12] J. Matoušek, On the l_2-discrepancy for anchored boxes, Journal of Complexity 14 (1998) 527–556.

[13] A. Owen, Variance with alternative scramblings of digital nets, ACM Transactions on Modeling and Computer Simulation 13 (4) (2003) 363–378.

[14] G. Ökten, W. Eastman, Randomized Quasi-Monte Carlo Methods in Pricing Securities, Journal of Economic Dynamics and Control 28 (12) (2004) 2399–2426.

[15] S. C. Clewlow L., Implementing Derivatives Models, John Wiley & Sons, 1998.

[16] R. Caflisch, M. Morokoff, A. Owen, Valuation of Mortgage Backed Securities Using Brownian Bridges to Reduce Effective Dimension, Journal of Computational Finance 1 (1) (1997) 27–46.

[17] A. Papageorgiou, The Brownian bridge does not offer a consistent advantage in Quasi-Monte Carlo integration, Journal of Complexity 18 (2002) 171–186.